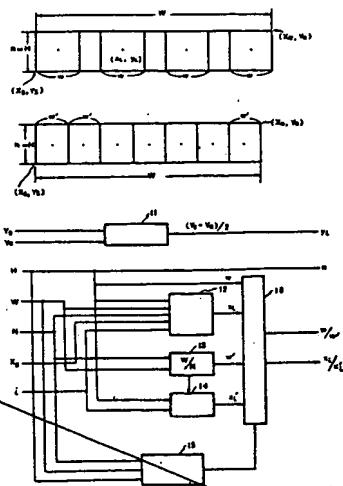


(54) CHARACTER REGION ARRAYING SYSTEM

(11) 59-47639 (A) (43) 17.3.1984 (19) JP
 (21) Appl. No. 57-158424 (22) 11.9.1982
 (71) FUJITSU K.K. (72) OSAMU KATOU(1)
 (51) Int. Cl³. G06F3/14, G06F15/20, G09G1/00

PURPOSE: To array character strings in a designated oblong region at accurate positions and with good balance, by calculating automatically the center coordinates of each character after discriminating whether or not the width or height of the character should be set smaller than the prescribed value.

CONSTITUTION: For input parameters, position coordinates Y_s , Y_e of an oblong region, height H , width W and input character number N , the position coordinates X_s of the oblong region, and the character order (i) are supplied. The coordinates Y_s and Y_e are fed to a center (y) coordinate calculating circuit 11 to calculate and then deliver each center coordinate Y_i . Then a center (x) coordinate ($W/N > H$) calculating circuit 12 feeds the character width $w = H, W, H, X_s$, i in the case of $W/N > H$ to obtain the center X coordinates X_i of an optional character frame and then transfers (w) and X_i to an output selecting circuit 16. In the case of $W/N < H$, a W/N calculating circuit 13 feeds W and N to obtain the character width (w'). A center X coordinate ($W/N < H$) calculating circuit 14 feeds W/N , H and (i) to obtain the center coordinates (x') of each character and then transfers X'_i and w'_i to the circuit 16. The circuit 16 selects and delivers either one of combinations (w) and X_i and w'_i and X'_i in accordance with the result of discrimination of a $W/N > H$ discriminating circuit 15.

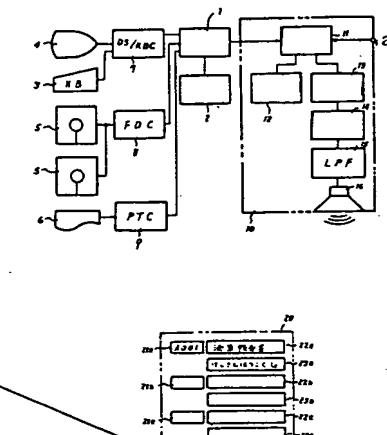


(54) WORD PROCESSOR PROVIDED WITH SOUND INDEX FUNCTION

(11) 59-47640 (A) (43) 17.3.1984 (19) JP
 (21) Appl. No. 57-157514 (22) 10.9.1982
 (71) FUJITSU K.K. (72) MAKOTO SUEDA
 (51) Int. Cl³. G06F3/16, G06F15/38, G11B27/10

PURPOSE: To improve the efficiency of retrieval, by adding a sound output function to the display of a display device to retrieve the documents which are stored in a word processor using the KANA (Japanese syllabary)/KANJI (Chinese character) conversion system.

CONSTITUTION: The document data produced by a word processor are stored in a floppy disk 5, and a document index list 20 stores the codes of the document numbers 21a..., titles 22a... written in KANA and KANJI and titles 23a... written in KANA, respectively. A control part 1 retrieves the list 20 in response to an input operation of a keyboard KB and displays successively the number 21a, the title 22a, the number 21b, the title 22b and so on on a screen of a display device 4. At the same time, the number 21a, titles 23a..., the number 21b and the title 23b are sent to a sound synthesis/output part 10. A sound synthesis control part 11 of the part 10 performs the sound synthesis 13 from a sound parameter memory 12 in accordance with the fed number numeric characters and KANA character codes as well as the monosyllabic editing or rule synthesis. The synthesized sounds are delivered through an amplifier 14, an LPF15 and a speaker 16 and synchronously with the display 4. Furthermore, the sound output speed can be controlled to an optional level.



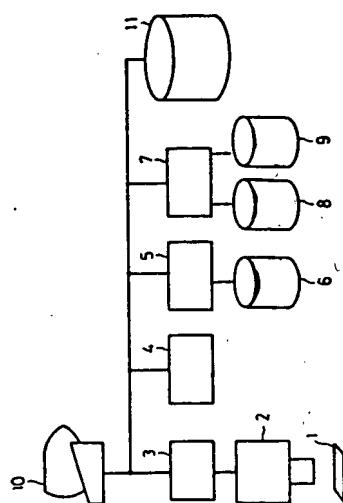
1: control part, 2: storage part, 13: sound synthesizer, 14: amplifier, 22a, 23a: report of accounts, a: speed indication

(54) PRODUCER OF VISITING CARD DATA BASE

(11) 59-47641 (A) (43) 17.3.1984 (19) JP
 (21) Appl. No. 57-156756 (22) 10.9.1982
 (71) NIPPON DENSHIN DENWA KOSHA (72) KUNIO OONO(3)
 (51) Int. Cl³. G06F7/22, G06F15/40

PURPOSE: To facilitate an input to a data base, by reading the characters printed on a visiting card and then sorting and rearranging automatically various types of character strings of the full name, title, address, telephone number, the post, etc.

CONSTITUTION: A visiting card 1 is read by a scanner 2 having the ≥ 40 pieces/mm reading resolution and then stored in a picture buffer 3 of $\geq 8M$ -byte capacity in the form of pattern information. The monochromatic picture information of the output of the buffer 3 is extracted out of a print region through a picture processing part 4. Then the largest character string is recognized as a full name, and a smaller character string printed above the full name is discriminated as a title. As the same time, the continuous numeric characters 9-11 including a hyphen is discriminated as a telephone number. Then, a code converting part 5 recognizes all characters by means of a printed KANJI (Chinese character) discriminating dictionary 6. The coded character string is discriminated as an address by means of a metropolis district discriminating dictionary 8 through a data sorting part 7. Then a post in a corporation is discriminated by a post discriminating word dictionary 9. The full name, title, address, telephone number, post, etc. are sorted and rearranged and then stored in a visiting card storage part 11.



⑯ 日本国特許庁 (JP) ⑯ 特許出願公開
 ⑰ 公開特許公報 (A) 昭59-47641

⑯ Int. Cl.³
 G 06 F 7/22
 15/40

識別記号

庁内整理番号
 7313-5B
 Z 6913-5B

⑯ 公開 昭和59年(1984)3月17日
 発明の数 1
 審査請求 未請求

(全 6 頁)

⑯ 名刺データベース作成装置

⑯ 特 願 昭57-156756

⑯ 出 願 昭57(1982)9月10日

⑯ 発明者 大野邦夫

横須賀市武1丁目2356番地日本
 電信電話公社横須賀電気通信研
 究所内

⑯ 発明者 高野陸男

横須賀市武1丁目2356番地日本
 電信電話公社横須賀電気通信研
 究所内

究所内

⑯ 発明者 深谷健一

横須賀市武1丁目2356番地日本
 電信電話公社横須賀電気通信研
 究所内

⑯ 発明者 長谷雅彦

横須賀市武1丁目2356番地日本
 電信電話公社横須賀電気通信研
 究所内

⑯ 出願人 日本電信電話公社

⑯ 代理人 弁理士 小林将高 外1名

明細書

1. 発明の名称

名刺データベース作成装置

2. 特許請求の範囲

(1) 名刺の記載事項を40本/mm²以上の解像度で読み取り可能なスキャナと、8メガバイト以上の容量を有する画像パックアンド、白黒画像情報を1画素ごとに操作可能な画像処理部と、印刷漢字識別辞書を有し文字パターン情報を文字コードに変換するコード変換部と、都道府県名識別辞書および所属識別用語辞書を有し文字コードデータを姓名、肩書、住所、電話番号、所属組織等の項目に整理分類するデータ分類部と、このデータ分類部からの分類データを検索管理し易い形に配列記憶する名刺データ記憶部とから構成されたことを特徴とする名刺データベース作成装置。

(2) 画像処理部は、名刺に印刷された文字列の最も大きな印刷文字列を姓名であると識別し、データベースに入力することを特徴とする特許請求の範囲第(1)項記載の名刺データベース作成装置。

(3) 画像処理部は、名刺に印刷された文字列において、最も大きな印刷文字列の前に、小さな印刷文字列が存在する場合、前記文字列を削除であると識別し、データベースに入力することを特徴とする特許請求の範囲第(1)項記載の名刺データベース作成装置。

(4) 画像処理部は、名刺に印刷された文字列において、括弧、ハイフンを含む9個以上11個以下の連続した数字が含まれ、かつその先頭の数字が“0”である場合、その文字列を電話番号であると識別し、データベースに入力することを特徴とする特許請求の範囲第(1)項記載の名刺データベース作成装置。

3. 発明の詳細な説明

この発明は、名刺の記載事項を読み取り、データベース化を行う名刺データベース作成装置に関するものである。

名刺の管理は、枚数が多いと煩雑なため名刺の内容を計算機に入力し、データベースに登録することにより、検索、管理等が行われている。しか

しながら従来は、けん盤からデータを手操作入力せねばならず、この入力作業が煩雑であるという欠点があつた。

この発明は、これらの欠点を除去するため、名刺を自動的に読み取り、計算機に入力し、データベース化するようにしたものである。以下この説明を図面について説明する。

第1図はこの発明の一実施例を示す構成図である。この図において、1は名刺、2はスキヤナ、3は画像バッファ、4は画像処理部、5はコード変換部、6は印刷漢字識別辞書、7はデータ分類部、8は都道府県識別辞書、9は所属識別用語辞書、10はキーボードディスプレイ、11は名刺データ記録部である。

これを動作するには、名刺1をスキヤナ2によってバターン情報をとして入力し、大容量バターンである画像バッファ3に蓄積する。スキヤナ2の読み取り分解能は、40本/mmぐらいは必要である。すなわち、名刺1に印刷される文字のうち、住所等で使用される最小のものは、約5ポイント

長に使うものであり、そのため、文字列10～11部分において鋭いピークを持ち、他の余白部分はバツクグラウンドノイズとなる。バツクグラウンドノイズレベル以上にスライスレベルL₁を設定すれば、文字列が文字列を構成する文字の組で抽出される。通常は、この第1段階の操作により、文字列が抽出される。

第2段階として、各文字列を前述の第1段階の走査方向に対し直角方向にバッファメモリ上で走査し、第1段階と同様にして頻度分布を求める。その代表例を第3図に示す。

文字列と直角方向の頻度分布においてもスライスレベルL₂を設定し、そのレベルと文字列の直なつた部分を列から分離し抽出する。その結果、ほぼ文字部分が抽出されるが、上下に分かれた文字（例えば「二」、「前」、「員」等）は、同一の文字でありながら分離して抽出されてしまう。そこで、次の第3段階として印刷されている漢字は、ほぼ正方形であるということを利用し、抽出された文字部分を補正し、1文字とする処理を行

活字程度（1.8mm×1.8mm）であり、それを64×64調素程度のドットマトリクスで読み取ることが必要だからである。

名刺1は白黒情報であるから、各画素は本体ならば1ビットでよいのであるが、印刷の線淡、紙の種類等のばらつきを考慮すると、1バイト程度必要となる。また、名刺1の大きさは、約5.5mm×9.0mm程度であり、40本/mmで画面に分解すると、1枚の名刺1のバターン情報は約8メガバイトになる。したがつて、画像バッファ3は、バターン情報のバッファエリアとして8メガバイト必要となるが、磁気ディスク（小形ウインチエストディスク）で十分間に合う値である。

次に、画像処理部4において、印字領域の抽出を行う。これは3つの段階に分類される。まず、第1段階は、特定の密度について名刺1の長辺方向をバッファメモリ上で走査し、短辺方向に対する頻度分布を求める。その代表的な結果を名刺1と対応させて第2図に示す。名刺1はほとんど横書の場合は紙面を縦長に、横書の場合は紙面を横

う。第4図にその一例を示す。

次に、性名部分の抽出を行う。これは、最もサイズの大きな文字部分を抽出することにより行われる。この際、肩書き部分の分離も同時に行われる。その状況を第5図に示す。すなわち、まず、文字列に対して直角方向に走査を行い第5図(a)のように文字の高さの差を検出し、高さの大きい文字と小さい文字とを区別し、次に第5図(b)のように文字列の方向に走査すると小さい文字。すなわち肩書き部分が姓名部分から分離される。

以上の処理で、各文字に対応した文字枠が抽出され、各文字枠に沿つて文字認識を行うことが可能となる。

次に、コード変換部5において、印刷漢字識別辞書6を用い文字認識処理を行う。通常はこの段階で、すべての文字が認識され、コード化される。この段階で、文字列のうちの姓名および姓名の上の小さな文字で印刷された肩書きは分離可能である。

次に、データ分類部7において、識別されコード化された文字列の分類を行う。住所の判定には、

都道府県名識別辞書⁸を用いる。この辞書には、全国都道府県および主要都市名が登録されており、文字列照合を行うことにより住所の文字列を識別する。なお、都道府県名識別辞書⁸には、全国市町村名まで登録されていれば、より正確な識別が可能になる。住所が2カ所記載されている名刺もあるが、これについても辞書を用いることにより、2カ所として同定される。

通常、電話番号は、住所の後の「電話」または「TEL」という文字の後に印刷されている。また、電話番号は、最初に「0」ではじまる市外局番と、市内局番および4けたの加入者番号から構成され、全けた数は、9～11けたである。以上の条件を用い、電話番号を同定することが可能である。

次の段階として所属識別を行う。通常、名刺¹の構成は第6図(a)、(b)のようなものが多い。この図において、Aは姓名、Bは肩書、Cは住所、Dは電話番号、Eは所属を示している。もしこの例に該当する文字列構成の名刺であるならば、こ

記述される。

なお、以上は一応横書きの場合を想定して書いたが、横書きの場合も全く同様となる。第6図(a)～(c)に対応する横書きの場合の構式は、第8図(a)～(c)のようになり、横書きにおいてもこのような構式が非常に多い。

社章等の図形が印刷されている場合、これ等を自動的に判別し、抽出することはかなり難しい。そのため、図形情報については、キーボードディスプレイ10によりオペレータ操作で事前にマスクするのが現実的な方法である。

O C R辞書で同定できない漢字あるいは隸つて認識された文字も、分類整理する以前にオペレータにより入力あるいは訂正されることが望ましい。以上2点のオペレータ操作は、第7図においてキーボードディスプレイ10と破線で示してある。

以上分類されたデータは、名刺データ記憶部11において検索管理し易い形に配列・記憶される。このような構造になつてるので、一般的の名刺をスキャナ入力するだけで名刺データベースを作成

の段階以前にA～Dの文字列は、各々姓名、肩書、住所、電話番号ということで同定される。したがつて、残存する文字列Eが唯一の残存文字列であれば、これは所属を示すと推定される。また、所属の文字列には、一般的の企業、官庁、各種機関の名称として用いられる用語(例えば「…株式会社」、「…大学」、「…部」、「…課」等)が含まれているのが普通であり、所属識別用語辞書⁹によりこれらの用語がマッチすれば、より高い確率で同定することが可能である。さらに、複数の文字列については、「その他データ」として分類される。

なお、肩書Eが第6図(c)のように姓名Aの上部にない場合もあり得るので、そのような場合のために肩書修有の用語(例えば「…課長」、「…部長」、「…博士」等)を辞書として用意し、マッチングをとり、一致すれば肩書レコードにする方法も考えられる。以上のプロセスをまとめると第7図のようになる。

第7図で、Iは知識ベース、IIは推論を含む処理手順、IIIは推定結果で、コードとしてメモリに

することができる。

以上説明したように、この説明は名刺に印刷された文字を読み取り、データベース化に必要な、姓名、肩書、住所、電話番号、所属等の各個文字列を自動的に分類し整理するので、容易にデータベースに入力でき、名刺の管理が容易になるという利点がある。特に、大企業幹部の名刺、代理士紹介等大量の名刺管理を必要とする人にとつては有効である。

4. 図面の簡単な説明

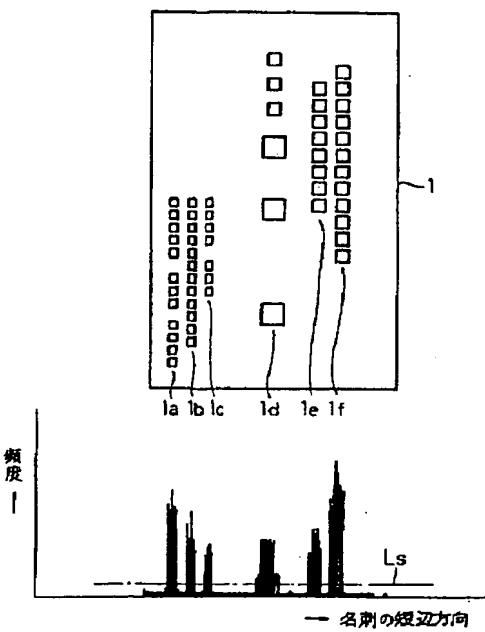
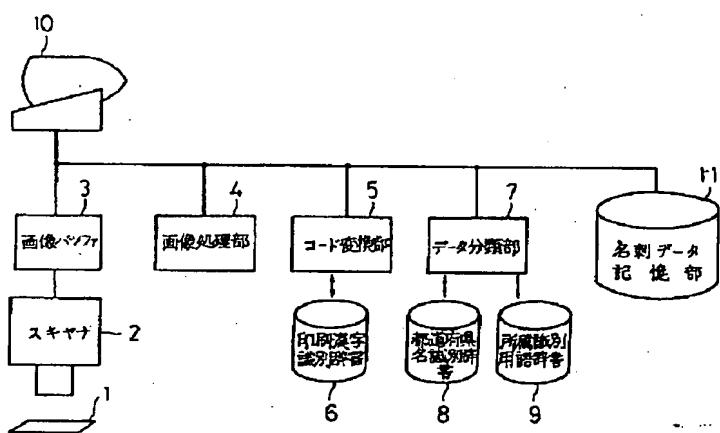
第1図はこの説明の一実施例を示す構成図、第2図は文字列を抽出する手法の原理図、第3図は文字を切り出すアルゴリズムを示す原理図、第4図は上下に分離した構造の文字を補正し、文字格を与えることを意味する原理図、第5図は姓名および肩書を分離し、文字格を適確に与えることを示す原理図、第6図(a)～(c)は横書きの名刺の代表的構造例を示す図、第7図はこの説明の実施例の処理手順を示す図、第8図(a)～(c)は横書きの名刺の代表的構造例を示す図である。

图中、1は名刺、2はスキャナ、3は画像バッファ、4は画像処理部、5はコード変換部、6は印刷漢字識別辞書、7はデータ分類部、8は都道府県名識別辞書、9は所轄識別用語辞書、10はキーボードディスプレイ、11は名刺データ記憶部である。

代理人 小林 栄一
島田松江(ほか1名)
近藤

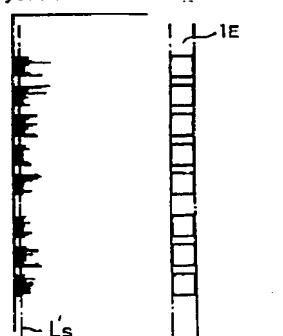
第 2 図

第 1 図



第3図

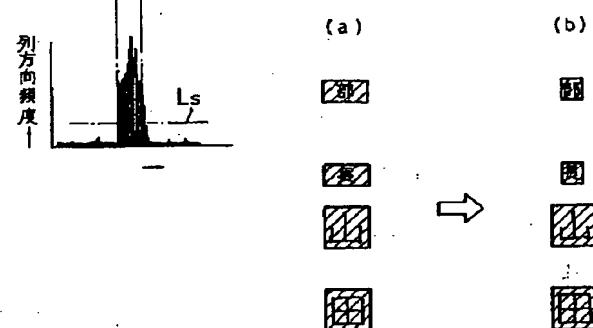
文字列と直角方向頻度



第4図

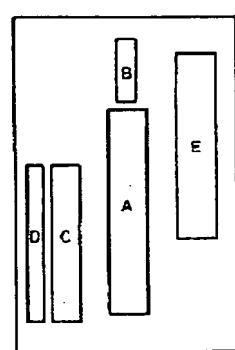


第5図

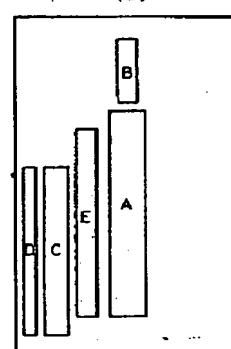


第6図

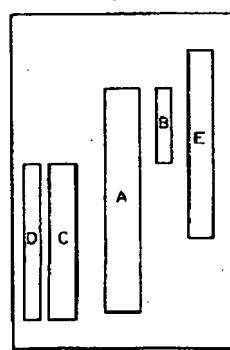
(a)



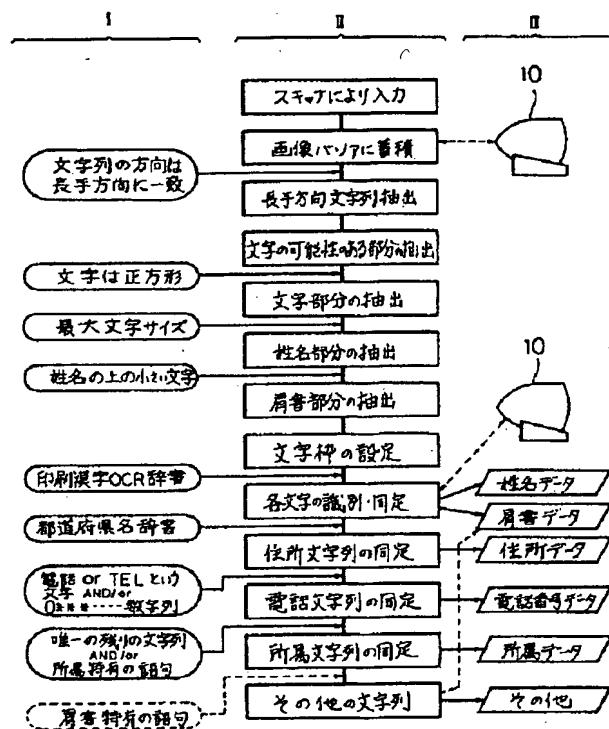
(b)



(c)

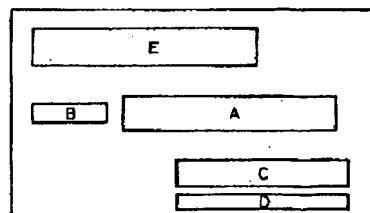


第 7 図

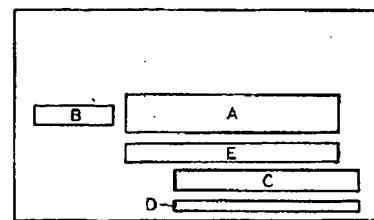


第 8 図

(a)



(b)



(c)

